已投稿会议：2024 数智时代知识管理与服务理论、方法、战略学术研讨会

在线心理社区提问用户信息需求主题挖掘
——以壹心理平台为例

王康龙 陈俊冶 王涟 殷彩明 吕艳华*

（山西医科大学管理学院　　山西 030600）

摘 要

[目的/意义]为优化在线心理社区信息资源的精细管理与高效利用,对用户信息需求的深层次挖掘并以此构建用户信息需求特征模型。[方法/过程]从信息需求认知层的主题维度对中文语境下在线心理社区内用户提问信息实施内容分析与统计研究,构建用户心理健康信息需求主题模型并对用户的需求特点进行归纳。[结果/结论]结果显示，用户心理健康信息需求的主题包括二次婚姻、未来规划、妯娌关系、婚姻生活、心理疾病史等五大类。[创新性/价值]研究的主要贡献在于:突破了被研究者心理健康需求研究过程中以问卷调查为主导的传统范式，提出了一种基于文本主题内容挖掘的新方案;文章致力于构建线上心理社区的理论模型，并在信息交互设计及功能提升方面提供改进措施以优化心理健康社区的系统结构。

关键词：在线心理社区；心理健康；用户信息需求；壹心理；内容分析

随着信息化进程不断推进，生活节奏变快，社会竞争严重，国民承受的压力逐渐增加。《中国国民心理健康发展报告》显示，心理健康已经成为比身体健康更加严峻的问题。心理障碍严重会演变为心理疾病，甚至可能造成悲剧发生。此外，与发达国家相比，我国的心理服务市场如今还不够成熟。但有关数据显示，我国居民对于心理和精神层面的需求正在逐年递增，消费意识的转变也不断推动精神类电商市场的发展。

心理咨询是指为全体公民在日常生活中可能面临的一般或严重心理问题而提供的专业心理咨询与干预服务，早先因医种成型晚出现了心理咨询资源短缺和分布不平衡的现象，但近年随着互联网＋时代的到来，在线心理咨询的方式逐渐流行，借助技术手段实现了医患的跨时空联系，大大节省了患者寻医的时间和成本。

国内心理健康领域有着壹心理、简单心理、松鼠社区等诸多社区，专业人员的入驻保证了信息质量与医疗服务的可靠性，但随着用户量的不断增加，部分社区用户信息量不断庞大且信息需求繁杂，使得社区信息资源管理的质量下降，用户反馈体验差的问题，这需要社区系统的整理归纳用户的信息需求，本文通过分析在线心理社区中用户群体发布的问题，对用

[通讯作者] 吕艳华，副教授，发表论文 50 余篇。

户信息的需求特点进行挖掘与归纳，并从系统信息资源组织、人机交互和功能优化等方面对在线心理社区构建与信息服务等方面提出建议。

1 相关研究综述

在国际学术界，领域研究可针对在线心理健康社区的信息、用户与社区三大层面展开。在用户生成内容方面，Park[1]基于 Reddit 论坛用户的讨论，应用多元文本挖掘法、定性分析技术，及可视化分析，剖析了焦虑症、抑郁症及创伤后应激障碍三类疾病的特点及共性。Feldhege[2]等人研发的程序，更为精确地收集和整理了该类论坛涉及的抑郁症相关话题，并统计了用户生产与个体特性的关联。Song[3]团队通过韩国公众对社交媒体发布自杀主题帖的细致分析，揭示了教育压力可能是引发青少年自杀的关键因素。类似，Pavlov[4]运用脚本处理 Twitter 数据，借助主题模型构建、情感算法分析，以及面板回归分析，探究文化因素对心理健康言论的影响。另外，韩潇[5]运用文本分析法、参与式观察法及深度访谈法等几种手段，就网络知识平台如何向精神病患者提供社区支持进行了深入研究。

基于深度研究用户体验行为，Liu[6]等人依托需求归属理论和社会支援理论展开线上调研，揭示出用户心理状态对在线健康社区四大核心布局（信息分享、协作、评价、倡议）存在负面影响。而 Zhou[7]等学者依托动机理论假设，采用壹心理网站数据进行可靠性评估，最后明确了内外动力的引导作用，有助于提升在线心理健康社区从业者的责任意识。此外，Oh[8]等人通过使用滚动式招募策略收集志愿者日常经历，揭示了支持性互动对社区认同感及生活满意度的深远影响。而王亭亭[9]则通过全面运用满足使用理论，采用半结构化访问法来洞察用户在使用心理咨询应用程序过程中的信息获取行为。曹南燕[10]等人则立足于在线信誉理论和科技采纳模型，借助问卷调查探寻影响用户持续使用意愿的关键因素——壹心理 APP 社区群组的归属感。

可以发现，当前针对用户心理健康信息需求的相关研究，能够在一定程度上揭示研究用户的信息需求类别。但从研究角度上，讨论用户心理健康信息需求时并未对信息需求主题进行分类概括[11, 12]。从研究方法上看，现有研究多以小规模的问卷调查和质性访谈为主，受研究设计和调查对象的影响，不同学者之间对用户心理健康信息需求的分类和排序存在差异，结论各有区分。因此，有必要结合内容分析与数据挖掘等技术手段对用户心理健康信息需求开展更加客观的研究。

2 实验设计

2.1 数据采集

在数据采集之前，对市面上在线心理健康社区调研分析，最终选取壹心理平台作为本研究实证对象，壹心理成立于 2011 年 7 月，是专业的心理服务提供平台，官方资料显示，截止 2020 年 12 月，拥有注册用户 2200 万，覆盖 125 个国家和地区；平台甄选 500 多名心理咨询师（入驻通过率低于 2%），庞大的用户规模和优质的医师资源使其成为国内心理社区日访问量最大的社区之一。利用 python 爬虫程序爬取该网站近三年所有提问数据（数据截止至 2023 年 7 月 1 日）共计 6584 条，经过人工数据清洗，去除重复、无用提问后保留有效数据 6117 条。

2.2 主题挖掘

运用 LDA 主题模型分析提问数据的潜在主题数量，再分词过程中，为改善词袋模型的词频问题，笔者利用 TFIDF 和 TEXTRANK 分别进行词频统计分析，经对比后最终选取 TFIDF 以改善词袋模型；在分词过程中剔除如“觉得”、“认为”、“岁”等无意义词汇，并将“双相情感障碍”、“妯娌关系”、“咨询师”等专业名词添入字典提升分词准确性[13]。通过 python 中 SK-learn 库构建模型，选取主题一致性衡量主题分类效果，据图 1 可知，当主题数为 12 时，coherence 值最大。

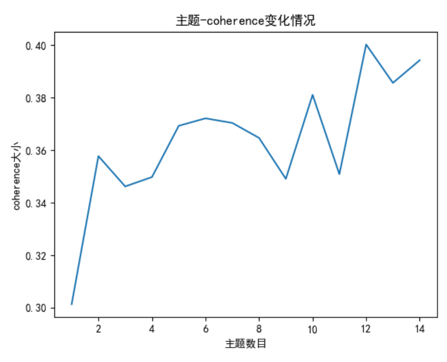


图 1 主题一致性曲线

设定主题数为 12，运行模型得出每个主题的主要内容及数据对应的主题概率（见表 1）。

表 1 主题对应内容

主题	对应主题内容
主题 1	亲戚感情问题
主题 2	父子/女感情问题
主题 3	恋爱感情问题
主题 4	婚姻生活问题
主题 5	个人生活问题
主题 6	心理疾病史问题
主题 7	妯娌关系问题

主题 8	夫妻家暴问题
主题 9	未来规划问题
主题 10	家庭子女问题
主题 11	青春期情绪问题
主题 12	二次婚姻问题

2.3 需求类别聚类

根据主题提取的结果，获取每条数据在 12 种主题中的概率，以 12 维概率向量表示，通过余弦相似度计算概率向量之间的距离从而得到用户间相似度。后根据用户间相似度进行降维，继而运用降维数据实施 K-means 聚类，使用手肘法和碎石图选定最优聚类数。由图 2 可看出当 K>5 时，SSE 的下降幅度放缓，整图在此形成“手肘”状，因次确定 K=5 时为最优聚类效果。数据经 K-means 聚类后得出问诊数据分组，将结果进行可视化展示（见图 3）。

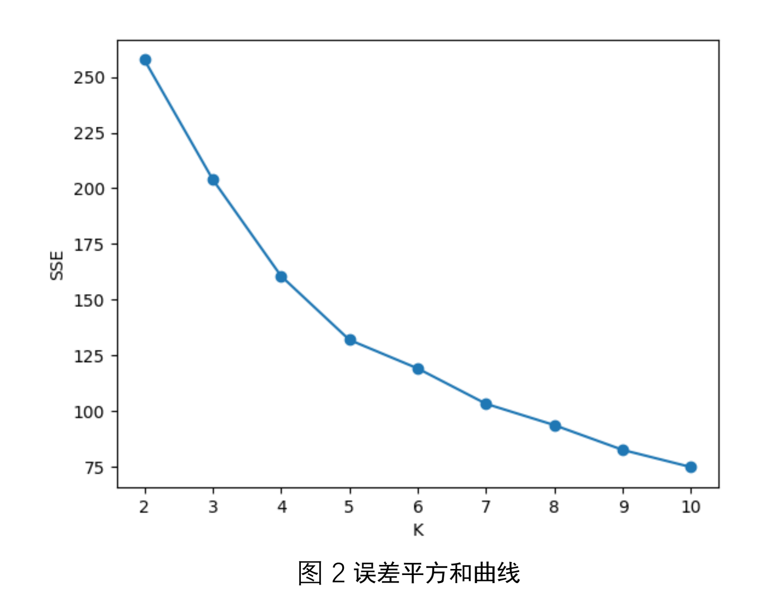


图 2 误差平方和曲线

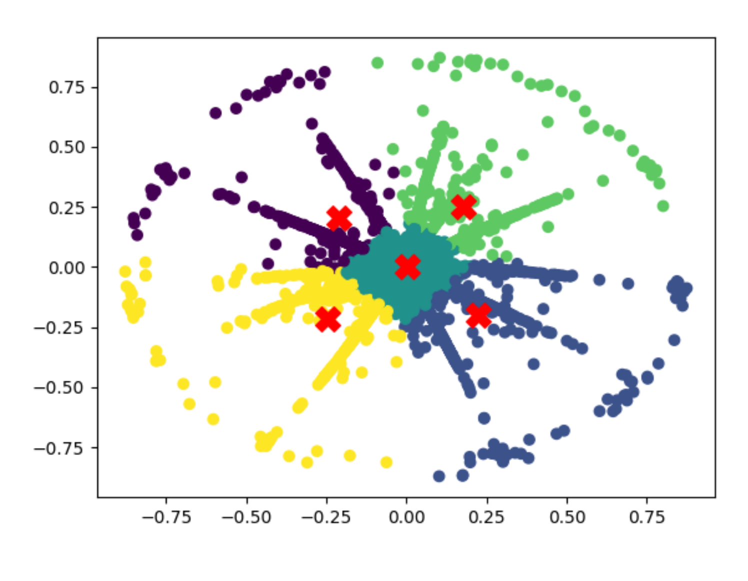


图 3K-means 聚类结果

3 实验结果

根据上述分析利用均值法算出每个簇的中心点和对应用户，将簇的中心用户对应主题作为该类的代表性主题（见表 2）。

表 2 簇团对应主题概率

主题 \ 中心点	簇团 1	簇团 2	簇团 3	簇团 4	簇团 5
	用户 1947	用户 5936	用户 89	用户 4311	用户 2763
主题 1	0.01894729	0.01202025	0.04166667	0.0783666	0.01094596
主题 2	0.01894731	0.01202028	0.04166667	0.01562494	0.19336259
主题 3	0.10352981	0.01202049	0.04166667	0.01562492	0.01094591
主题 4	0.0189473	0.01202007	0.04166667	0.33571085	0.01094615
主题 5	0.01894741	0.0120203	0.04166667	0.01562487	0.01094883
主题 6	0.0189532	0.01202081	0.04166667	0.01562568	0.36119558
主题 7	0.3588164	0.3856747	0.54166667	0.33422893	0.34691328
主题 8	0.01894738	0.01202004	0.04166667	0.01562486	0.01094621
主题 9	0.01894883	0.39502201	0.04166667	0.01562519	0.01094597
主题 10	0.01894796	0.0120205	0.04166667	0.01562512	0.0109462
主题 11	0.01894964	0.01202008	0.04166667	0.01562601	0.0109475
主题 12	0.36711748	0.11112048	0.04166667	0.12669204	0.01095583

研究数据中在线用户提问数据包含隐含主题 12 个，经降维聚类后共总结出二次婚姻(1.90%)、未来规划(1.52%)、妯娌关系(1.18%)、婚姻生活(1.34%)、心理疾病史(84.31%)共五个大类。

4 讨论与建议

4.1 在线用户心理健康信息需求类型及特征讨论

对上述聚类结果进行分析讨论，笔者采用 python 中的 WordCloud 包以绘制词云的方式将主题内容进行可视化展示。

(一) 二次婚姻生活纠纷相关问题

关于二次婚姻的提问，多是用户在婚姻中存在了不信任对方的情绪，在遇到某些事情时产生分歧，造成对伴侣的不信任，从图 5 可以看出，双方的矛盾点主要集中在对前段婚姻的阴影、继养的子女、夫妻生活、男方不正当行为（偷拍）等方面。另外结合对爬取数据的时间统计分析折线图（见图 6）得出，该类提问的频次正逐年上升，因此平台可对此多加关注，并以此作为招聘咨询师的依据。



图 5 二次婚姻生活纠纷词云

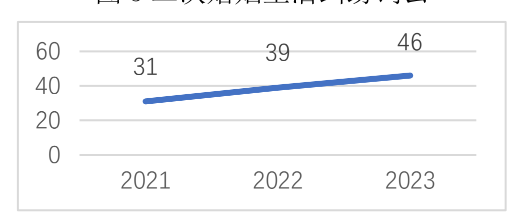


图 6 二次婚姻统计折线图

(二) 未来规划迷茫相关问题

在线用户在人生面临选择或者不满足生活现状时，常伴有消极情绪的出现，易产生焦虑症或抑郁症倾向，结合图 7 和内容可以看出，该类用户的信息需求一般是对某事物的选择出现了两难（定居、调解、解决办法），或是对个人生活状态出现不满从而引发情绪（拖延、放纵、懊恼）。



图 7 未来迷茫词云

(三) 妯娌关系不和等问题

妯娌关系不和是家庭中常见的问题之一，通常指的是夫妻的姐妹之间存在矛盾、冲突或不愉快的情况，据图 8 可看出家庭中的经济、子女教育、家务分配等可能加剧了妯娌关系的紧张与不和；另外妯娌间的缺乏沟通也使得矛盾不断积累，导致家庭氛围紧张，影响到了家庭成员的情绪；此外，该类数据的提问量也是上升趋势，且 2023 年上升尤为明显（见图 9）。



图 8 妯娌不和词云

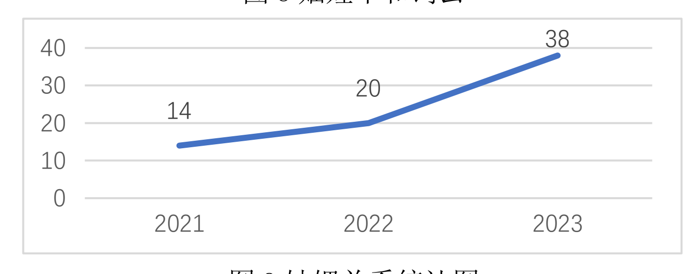


图 9 妯娌关系统计图

(四) 婚姻生活不幸福等问题

该类用户的信息需求主要是和自己婚姻的另一半生活不幸福，时常感到压抑因而想寻找解决办法，通过图 10 分析可得出用户婚姻间矛盾主要集中在子女抚养（生小孩）、家庭开支

(保姆)、一方疑似出轨（变态、网上聊天）等问题，另外通过审阅原数据了解到该类用户的情绪障碍已经严重到一定程度，部分已经患上忧郁症、焦虑症等情感疾病。



图 10 婚姻生活词云

(五) 心理疾病史相关问题

心理疾病史问题对应的用户大多患有心理疾病或有过疾病史，该类的信息需求种类较多，图 11 可看出“心理疾病”权重最高，其他只显示了部分社交属性词，查阅原数据后发现，用户部分患有 ADHD（注意力缺陷多动障碍）、PSTD（创伤后应激障碍）、GAD（广泛性焦虑障碍）、MDD（抑郁症）等疾病，提出的问题也是因疾病原因遇到的情感障碍，这类用户在平台提问数据中占比最高，且常年活跃，是平台的重点服务对象。



图 11 心理疾病问题词云

4.2 建议

平台的功能设计最终都会把用户作为中心服务对象，基于用户的设计是当下系统开发设计领域的基本设计思维[14]。本文以在线心里健康平台用户的信息需求为中心，从交互设计、

功能模块两个方面为壹心理平台提出优化建议。

(一) 信息需求结构化

在现有的在线心理健康社区问答模块中，鲜少涉及系统化的交互设计。鉴于平台用户的教育背景各异，他们在描述信息需求时语言结构复杂且参差不齐[15, 16]。随着信息量的激增，需要利用机器自动识别用户所创建的内容及信息需求，然而文本挖掘对于语义理解及信息需求情境的提取仍然面临了巨大挑战。该文通过深入分析信息需求特性，发现用户的信息表达蕴含丰富的情境要素，如个人情况、提问动机和寻求的信息支持。因此，结合用户信息需求特点进行设计，指导用户开展结构化的信息表达，有助于更准确地辨识用户信息需求[17]。

(二) 功能设计优化

在线心理社区用户普遍存在对评价和推荐的信息需求，诸如心理健康历史主题下，用户对在线心理咨询师推荐的信息需求比重较大。然而，诸多提问存在相似性，且解答各异，使得用户难以从繁琐信息中得出明确结论[18, 19, 20]。对此，在线心理社区应正视这一需求，在功能设计方面加以改进，增设独立的推荐与评价模块；此外，针对当前仅采用文字问答模式接受用户付费提问的服务方式，平台可以考虑加入视频或音频模块，以特定情境下与提问方进行具有亲和力的语音通讯、视频交流等形式，进而减少用户咨询期间的焦虑情绪，提高心理治疗的有效性。

5 结语

本文以壹心理平台的用户提问数据为例，通过爬虫技术获取在线心理领域的用户数据，经过多个步骤处理，最终确定用户主要信息需求类别。这些研究结果有利于平台全面理解用户的需求与特征，据此改进平台系统设计，添加结构化信息输入，提高用户填写质量，减少语义理解歧义的情况，更好地满足用户需求，提升服务品质。当然，本研究也有其局限性：受制于数据存储和处理能力，研究仅选取单个平台，未考虑多个平台数据多源融合的情况，未来研究可补充更多数据源，拓宽分析范围。总体而言，文章对用户心理健康信息需求特点的分析能够为在线心理社区的交互和功能设计优化提供思路与建议，本篇论文不仅创新了再心理健康领域数据分析方法，还为在线心理社区平台在未来的深入研究提供有力支持。

参考文献

[1] Park A, Conway M, Chen A T. Examining thematic similarity, difference, and membership in three online mental health communities from reddit: A text mining and visualization approach[J]. Computers in Human Behavior. 2018, 78: 98-112.

[2] Feldhege J, Moessner M, Bauer S. Who says what? Content and participation characteristics in an online depression community[J]. Journal of Affective Disorders. 2020, 263: 521-527.

[3] Song J T M D. Data Mining of Web-Based Documents on Social Networking Sites That Included Suicide-Related Words Among Korean Adolescents[J]. The Journal of adolescent health: official publication of the Society for Adolescent Medicine. 2016, 59(6).

[4] Pavlova A, Berkers P. "Mental Health" as Defined by Twitter: Frames, Emotions, Stigma[J]. Health Communication. 2020.

[5] 韩潇. 网络知识平台对精神疾病患者的社会支持研究[J]. 安徽大学. 2019.Xiao Han. 2019.

[6] A S L, A W X, A C F, et al. Social support, belongingness, and value co-creation behaviors in online health communities - ScienceDirect[J]. Telematics and Informatics., 50.

[7] Zhou J, Zuo M, Ye C. Understanding the factors influencing health professionals' online voluntary behaviors: Evidence from YiXinLi, a Chinese online health community for mental health[J]. International journal of medical informatics. 2019, 130(Oct.): 103931-103939.

[8] A H J O, B E O, B L R. How does online social networking enhance life satisfaction? The relationships among online supportive interaction, affect, perceived social support, sense of community, and life satisfaction[J]. Computers in Human Behavior. 2014, 30(1): 69-78.

[9] 王亭亭，莫扬. 心理咨询类APP用户的信息获取行为研究——基于四款APP的沟通分析[J]. 科普研究. 2017, 12(6): 8.Tingting Wang，Yang Mo. 2017, 12(6): 8.

[10] 曹南燕，曾艳，张州艳，等. 心理类APP用户持续使用意愿的影响因素探析——以"壹心理"为例[J]. 统计与管理. 2020, 35(4): 6.

Nanyan Cao，Yan Ceng，Zhouyan Zhang，等. 2020, 35(4): 6.

[11] 赵安琪，赵海平，路培鑫. 基于社会化问答社区的抑郁症健康信息需求研究[J]. 中华医学图书情报杂志. 2018, 27(9): 8.Anqi Zhao，Haiping Zhao，Peixin Lu. 2018, 27(9): 8.

[12] 生荾荾. 当前我国信息用户的信息需求特征研究[J]. 技术与创新管理. 2011, 32(5): 4.Suisui Sheng. 2011, 32(5): 4.

[13] 赵安琪，赵海平，路培鑫. 基于社会化问答社区的抑郁症健康信息需求研究[J]. 中华医学图书情报杂志. 2018, 27(9): 8.Anqi Zhao，Haiping Zhao，Peixin Lu. 2018, 27(9): 8.

[14] 成全，邓婷燕. 在线母婴社区的用户健康信息需求挖掘——基于妈妈网的实证[J]. 现代情报. 2022, 42(5): 8.

Quan Cheng，Tingyan Deng. 2022, 42(5): 8.

[15] 徐一方. 消费者健康信息需求模型的构建与应用[D]. 华东师范大学, 2015.

Yifang Xu. 华东师范大学, 2015.

[16] 刘冰，历鑫，张赫钊，等. 网络健康社区中身份转换期女性信息需求主题特征及情感因素研究——以"妈妈网"中"备孕版块"为例[J]. 情报理论与实践. 2019(5): 6.Bing Liu，Xin Li，Hezhao Zhang，等. 2019(5): 6.

[17] 涂为. 公共图书馆老年读者信息需求特征及信息服务优化路径研究[J]. 河南图书馆学刊. 2023, 43(9): 40-43.Wei Tu. 2023, 43(9): 40-43.

[18] 徐一方. 消费者健康信息需求模型的构建与应用[D]. 华东师范大学, 2015.

Yifang Xu. 华东师范大学, 2015.

[19] A S L, A W X, A C F, et al. Social support, belongingness, and value co-creation behaviors in online health communities - ScienceDirect[J]. Telematics and Informatics., 50.

[20] 李重阳，翟姗姗，郑路. 网络健康社区信息需求特征测度——基于时间和主题视角的实证分析[C]. 2016.Zhongyang Li，Shanshan Zhai，Lu Zheng. 2016.